Journal of Informetrics 15 (2021) 101121

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Using graph embedding and machine learning to identify rebels on twitter

Muhammad Ali Masood, Rabeeh Ayaz Abbasi*

Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 2 March 2020
Received in revised form 7 November 2020
Accepted 25 November 2020
Available online 5 December 2020

Keywords:
Rebels
Social network analysis
User graph
Supervised Rebel Identification (SRI)
Machine learning

ABSTRACT

During the last two decades, the number of incidents from extremists have increased, so as the use of social media. Research suggests that extremists use social media for reaching their purposes like recruitment, fund raising, and propaganda. Limited research is available to identify rebel users on social media platforms. Therefore, we propose a Supervised Rebel Identification (SRI) framework to identify rebels on Twitter. The framework consists of a novel mechanism to structure the users' tweets into a directed user graph. This user graph links predicates (verbs) with the subject and object words to understand semantics of the underlying data. We convert the user graph into graph embedding to use these semantics within the machine learning algorithms. Apart from the user graph and its embedding, we propose fourteen other features belonging to tweets' contents and users' profiles. For evaluation, we present the first multicultural and multiregional dataset of rebels affiliated with nine rebel movements belonging to five countries. We evaluate the proposed SRI framework against two state-of-the-art baselines. The results show that the SRI framework outperforms the baselines with high accuracy.

1. Introduction

Rebels are individuals who exercise opposition to the government using force or armed resistance. They use social media to share their ideology. As, social media allows users to share a broad range of content, which encourages rebel users to disseminate propaganda with the aim of attracting and recruiting others to their cause to convince people. The aim of such is to gain funding, sympathies, and recruitment through social media. In the past, an extremist group has used its popularity in social media to recruit about 15,000 foreign fighters across the globe (Schneider, 2015).

The proposed research investigates key features and patterns in identifying rebel users on Twitter. These features and patterns not only identify rebel users, but they may be applicable in identifying global miscreant users. The proposed framework Supervised Rebel Identification (SRI) is the first of its kind, which can function as a benchmark for future research. Besides that, the proposed framework may help governments to promptly identify rebel users and resolve their grievances.

Researchers have not yet focused on developing a dataset of rebel users. As a result, this research aims at developing a dataset constituting **rebel**, **counter rebel**, and **normal** users. **Rebels** are the users who forcefully challenge the government entities, formally defined in Section 2. **Counter rebel users** are the normal users except that they share more tweets against

* Corresponding author.
E-mail address: rabbasi@qau.edu.pk (R.A. Abbasi).

https://doi.org/10.1016/j.joi.2020.101121
1751-1577/

the rebellious movements (defined in Section 2). Finally, the **normal users** are those who share tweets on any topic, as discussed in Section 2.

As the behavior/pattern of the rebel users tends to pertain to social aspects, cultural norms, laws, and region, we develop a multicultural and multiregional dataset of nine rebellious movements. These rebellious movements are picked using various criterion, which are discussed in detail in Section 4. The proposed multicultural and multiregional dataset allows the proposed model to identify generalized patterns on a diverse set of rebel users.

This research has the following contributions:

1. We present a new dataset having more than 284,000 tweets of both rebel, counter rebel, and normal users. The proposed dataset targets nine rebellious movements active in five countries: United Kingdom, Turkey, Syria, Iraq, and Pakistan.
2. We propose a novel user graph to structure the tweets in a way similar to an ontology. The proposed user graph highlights the ideology and stance of a user.
3. We propose fifteen features belonging to three categories (contents, user profile, and graph embedding features) to identify rebel users.
4. We analyze various aspects of rebel, counter rebel, and normal users.

2. Background and related work

Researchers among social sciences have extensively studied freedom movements and rebels. However, these topics are not widely studied in computer science because of the inadequate number of conferences and journals (J. Tinnes, 2020). Due to limited literature on rebel identification, this section also discusses techniques from radicalization. These techniques are helpful in developing a useful framework for identifying rebel users.

This research focuses on identifying rebels based on their activities on social media. Therefore, it is important to first define rebel groups. Users affiliated with these groups are considered rebels. Mueller (2014) define a rebel group as (Mueller, 2014, Page 10):

> *"A group that is armed, not under state control, and use force as part of their strategy to challenge the state. These groups act in opposition to the government or state in control of the territory, or, where there is no government to oppose, use force against other groups contending for power"* (Mueller, 2014, Page 10).

According to the definition, we exclude those groups that perform violent activities across the globe (a.k.a. terrorist groups). Terrorist groups have a global footprint, so they have more resources and impact. On the other hand, rebel groups have limited resources, and they are confined within a certain region. Therefore, patterns of terrorist groups may differ with the rebel groups.

In this research, we consider a person as a rebel if he either shows affiliation with the rebel groups or he forcefully opposes government laws and provokes the community to take revenge from the state. In particular, he shares propaganda, hatred, or threats to force the government to accept his demand of a sovereign state.

Apart from the rebel users, we have picked counter rebel users. These users use similar vocabulary, but the context is opposite to that of a rebel user. This overlap of words makes their identification task more challenging. In simple terms, the counter rebel users are normal persons, but they have an intention to criticize the rebellious movements. For example, a TV actor who actively criticizes rebellious movement (like Kurdistan movement and Welsh independence movement) can be considered as counter rebels. Counter rebel users are important when identifying rebels, since they might be classified as rebel users because of their overlapping vocabulary with the rebels. Apart from a counter rebel user, a normal user is neither a rebel nor a counter rebel. We added normal users to depict a real-world scenario.

2.1. Role of twitter in radicalization

The ease of using social media enables certain users to share hatred with local and global communities. According to analysis (on an anonymous social network named "Whispers"), about 18% of posts (whispers) are radical (Wang et al., 2014). Another analysis reveals that the average recruitment age of German fighters in ISIS was 27 years, whereas females radicalize at a much younger age of 15–18 years (Reynolds & Hafez, 2017). Youth involvement in the negative activities may result in an increase of violence and hatred.

Many miscreants use Twitter, about 45,000–90,000 ISIS accounts were suspended by Twitter, but many remained active (Rowe & Saif, 2016). The algorithm for the identification of these kinds of users is not publicly available. Besides that, this algorithm has inaccuracies. For example, Twitter once suspended the account of a normal user "Iyad El-Baghdadi", because it mistaken him to be the head of ISIS "Abu Bakr al-Baghdadi" (BBC, 2019).

2.2. Social network analysis

Social network analysis (SNA) uses graphs (also called networks) to study relations. Researchers have used Link Based Bootstrapping (LBB), which takes a seed user (e.g., a known extremist user) and parses his followers, hyperlinks, friends, and subscriptions to form a social network. Similarly, (Sureka, Kumaraguru, Goyal, & Chhabra, 2010) and (Chatfield, Reddick, &

Brajawidagda, 2015) have also used the Link Based Bootstrapping (LBB) to create a network of ISIS users. Apart from creating a network, the LBB approach can be used to generate a dataset, however, it might require manual labeling to discard irrelevant users.

The authors of (Ovelgonne, Kang, Sawant, & Subrahmanian, 2012) proposed the use of a covertness centrality measure that maps centrality scores and communication frequency of a new user with the known covert nodes (Ovelgonne et al., 2012). In the same manner, some researchers have proposed adding scores of the extremist neighborhood within the centrality measures (Gialampoukidis et al., 2017; Wadhwa & Bhatia, 2016; Wei & Singh, 2017). Similarly, authors in (Wei, Singh, & Martin, 2016) used user ties with the extremist group as a feature in the machine learning algorithms (Wei et al., 2016). Apart from this, authors in (Husslage, Borm, Burg, Hamers, & Lindelauf, 2015) and (Richey & Binz, 2015) rank extremists based on different attributes within their network. Another interesting aspect of social network is to create communities; therefore, Li et al. (2012) used topical structure to form community of normal users (Li et al., 2012).

Researchers proposed techniques that form a word graph using co-occurring words (Bordoloi & Biswas, 2019; Castillo, Cervantes, Vilarino, Báez, & Sánchez, 2015) to identify sentiments of a user. For example, the algorithm proposed by (Castillo et al., 2015) identifies the most representative words using the word graph. These words are used for sentiment classification. This technique uses co-occurrence to form a word graph, however, such graph may not identify relationships among words. For this sake, (Rousseau, Kiagias, & Vazirgiannis, 2015) proposed a technique to extract the core concepts within the co-occurrence graph. Even with the core concepts, the co-occurrence graph can only extract information within one entity, i.e., word. In real-world, sentiments of a user may depend on the relationship between multiple entities (like words-emojis and words-reviews). In graph theory, we can analyze the relationship among two disjoint sets of entities in a bipartite graph. Many researchers have used bipartite graphs to gain semantic context (F. Haneef, M.A. Sindhu, M.N. Noor, & A. Daud, 2020; Pan, Ni, Sun, Yang, & Chen, 2010; Seyednezhad, Fede, Herrera, & Menezes, 2018; Sindhwani & Melville, 2008).

2.3. Supervised machine learning

Many complex real-life problems are solved using supervised machine learning. In line with this, researchers propose to utilize linguistic rules for the classification of extremist users. Many researchers classify sentences by mapping them to a predefined list of radical or hateful words (Badawy & Ferrara, 2018; Bermingham, Conway, McInerney, O'Hare, & Smeaton, 2009; Chen et al., 2004; Fernandez, Asif, & Alani, 2018; Skillicorn, 2010; Sureka, 2012; Wadhwa & Bhatia, 2014, 2015). These predefined words often relate to a specific region or culture. To overcome this issue, researchers use contents (like bag of words (BOW), n-gram, or emotional words), user profile (including previous words, past activities, tweets count, or retweets count), and social media profile features (like verified account, profile picture, or geo-location) to classify extremist users (Alvari, Sarkar, & Shakarian, 2019; Debnath, Das, & Das, 2017; Hartung, Klinger, Schmidtke, & Vogel, 2017; Scanlon & Gerber, 2014; Wei & Singh, 2017). (Agarwal & Sureka, 2015) also used content features like war keywords, hashtags, religious terms, negative emotions, emotions, and slang to identify radical users. The authors concluded that the war keywords, offensive words, and religious terms are useful features for predicting radical users. Such a research is limited to the words used by ISIS users and may not be applicable in other scenarios.

In contrast to the use of predefined vocabulary features, researchers in other studies (Fernandez & Alani, 2018; Saif, Fernández, Rowe, & Alani, 2016) propose extraction of semantic concepts such as entity, entity types, and categories to identify the pro-ISIS stance of a user. These semantic concepts can also help in finding similar users. On the contrary, n-gram based features are used to classify radical contents (Choi, Ko, Kim, & Kim, 2014; Fernandez et al., 2018). In the same manner, authors in other investigations (Agarwal & Sureka, 2015; Wadhwa & Bhatia, 2013) have mapped the vectors of n-gram (of a new user) with known extremist users.

Many techniques discussed in this section use predefined rules or vocabulary to identify radical users. Even if we ignore the fact that rebel and radical users have different patterns; these predefined rules or vocabularies are not robust, thus ignore the changing demands for identifying rebel users. We have found that classification is not widely used for identifying rebels because it requires ground truth or manual labeling. Apart from labeling, sometimes extracting data of rebel users can become tedious, because their accounts are often suspended by Twitter.

3. Supervised Rebel Identification (SRI)

It is important to identify the context of a user before classifying him as a rebel. For this sake, we need to first identify key features that help in the identification of rebel users. Therefore, we propose the SRI framework, shown in Fig. 1. In this framework we first list various features from tweets and user profiles. Next, we use k-fold cross validation to pick the key features. Besides that, we combine distinctive features together to evaluate accuracy. We intend to pick those features or their combinations that achieve better accuracy. Finally, we select the best features and use them in machine learning algorithms to identify rebel users.

In this research, we proposed features belonging to three categories: contents, user profile, and graph embedding. The objective of content features is to use words, hashtags, bigrams, and past sentiments to identify rebel users. On the contrary, user profile features aim to utilize the user influence attributes in the machine learning algorithms. Finally, we propose to capture the semantics of a user by creating a user graph from the part of speech (POS) tags. We convert the graph into a

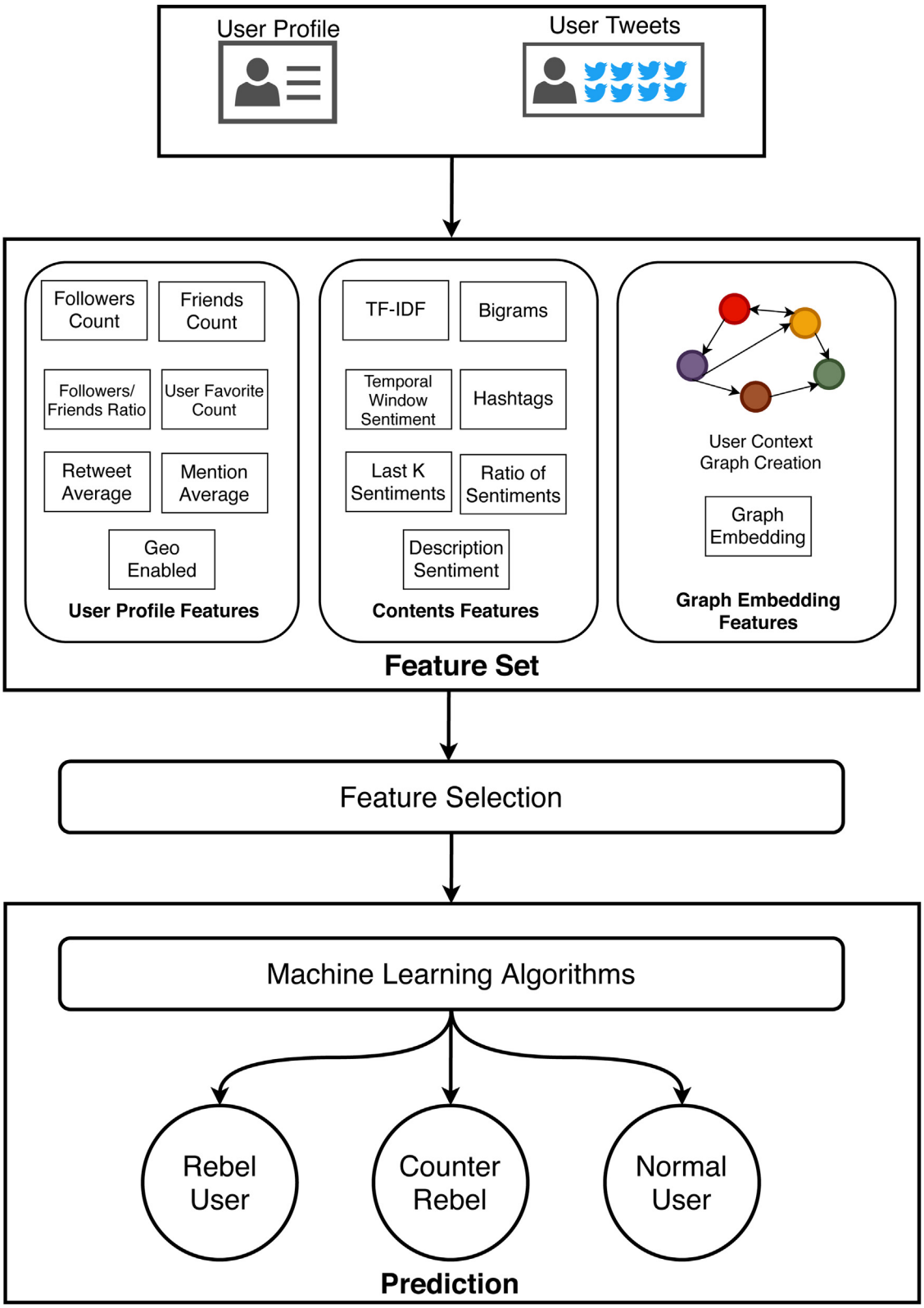

Fig. 1. The proposed SRI framework for identification of rebel users. The framework consists of three main categories of features, i.e., user profile, contents, and graph embedding features.

lower dimension feature vector using graph embedding. After listing all the features, we select the best features and their combinations based on results based on the wrapper method discussed in Section 3.4.

3.1. Content features

Contents are key indicators for identifying the inclination of a user towards rebellious groups. In the past, researchers have used content features to classify users or articles (Alvari et al., 2019; Debnath et al., 2017; Hakim, Erwin, Eng, Galinium, & Muliady, 2014; Hartung et al., 2017; Scanlon & Gerber, 2014; Wadhwa & Bhatia, 2013; Wei & Singh, 2017). In this research, we extract five features, i.e., term frequency-inverse document frequency (TF-IDF), bigrams, hashtags, sentiments in Twitter profile description, and sliding window-based sentiments features. The first feature is the term frequency-inverse document frequency (TF-IDF) feature, which is one of the well-known approaches for classification (Dadgar, Araghi, & Farahani, 2016; Hakim et al., 2014). In this research, we intend to use TF-IDF features to identify discriminant words of rebel, counter rebel, and normal users. The TF-IDF model is useful, but it ignores the relationship between two consecutive words. As a result, the second feature extracts these relationships using bigrams. In the past, bigrams have been used to identify relationships among words for classification (Agarwal & Sureka, 2015; Wadhwa & Bhatia, 2013; Wang & Manning, 2012), so we make use of the same philosophy to identify rebel and other users. The third feature captures the user semantics from the hashtags, these semantics are useful in solving various problems (Agarwal & Sureka, 2015; Nam, Lee, & Shin, 2015; Wang, Wei, Liu, Zhou, & Zhang, 2011). Therefore, we capture these user semantics to understand a user's perspective during classification. According to the analysis on the proposed dataset, rebel users share more hashtags to show their alliance with a certain cause. In this research, we also extract the same features listed above for the profile description. We aim to separate features of the profile description from the content features because features from the content can overwhelm the profile description features, thus neglecting important aspects within the profile description.

After obtaining the users' perspectives, the fourth feature assumes that the rebel users mostly accuse their governments in their profile descriptions. Thus, we extract sentiments separately from the profile description. We hypothesize that due to limited number of words in a profile description, it may result in inaccurate sentiment classification. Consequently, we emphasize on the integration of sentiments of the contents. Generally, opinions of users change over time, so it is important to obtain sentiments in different temporal sliding windows (Masood, Abbasi, & Wee Keong, 2020). We used three sliding window features: the first sliding window feature uses the temporal window to extract the user's recent sentiments. The temporal window feature may have few or no past tweets in a shorter time span. Therefore, the second sliding window feature (last k sentiments) takes past k tweets irrespective of time. Both sliding windows do not capture the overall sentiments of a user. As a result, we use the ratio of sentiments feature to obtain the ratio of the positive, neutral, and negative sentiments from all the past tweets (Masood et al., 2020).

3.2. User profile features

The rebel users intend to gain influence to convince and recruit individuals from across the globe. A Twitter profile contains numerous features that are used to identify hatred or radical users (Wei & Singh, 2017; Xu & Lu, 2016). In this research, we picked seven key features, namely followers count, friends count, the ratio of followers and friends, user favorite count, retweet count, the average number of mentions and Geo-enabled. The first feature captures the user's influence (by counting the number of followers) (Alvari et al., 2019). In this research, we hypothesize that rebel users are more dedicated to the cause, as a result, they may have a higher number of followers than other users. Similarly, the number of friends is also commonly used to identify user influence (Alvari et al., 2019). In this research, we hypothesize that rebel users will have a smaller number of friends because of the risk of affiliating with a rebel user. The ratio of followers and friends can approximate a user's popularity. We hypothesize that rebel users would have a high ratio of followers and friends because of their dedicated following. In the same manner, retweet is another indicator used to identify radical user (Debnath et al., 2017). In this research, we hypothesize that the dedicated following (of rebel users) may lead to a greater retweet count.

Mentions are the way to see the interactions of the users with other users. In the past, researchers have also used it for classification (Hartung et al., 2017; Xu & Lu, 2016). Rebel users intend to convey their agenda across different communities such as, journalists, human rights organizations, influential leaders, and the UN. This phenomenon led us to use the average number of mentions per tweet as a sixth feature. The last feature depicts whether the user has enabled geographical coordinates or not. Generally, rebel users do not enable the geo-coordinates of their tweets to remain hidden. Therefore, enabling of geo-location can be used as a feature to classify radical users.

Algorithm 1. User graph creation algorithm

INPUT: Tweets of user j $tweets_j$
OUTPUT: Graph $G_j = (N_j, E_j)$
  procedure CREATEGRAPH($tweets_j$)
    for $tweet_i \in tweets_j$ **do**
      Dependency=extractPOSDependency($tweet_i$)
      for $word_l$, $POStag$, $word_m \in$ Dependency **do**

$stemWord_l$=stemming($word_l$)
$stemWord_m$=stemming($word_m$)
if *POStag*!="compound" **then**
$N_j = N_j \cup stemWord_l \cup stemWord_m$
$E_j = E_j \cup (stemWord_l, stemWord_m)$
else
$compoundWord$=concatenate($stemWord_l$, $stemWord_m$)
$N_j = N_j \cup (compoundWord)$
d_1=extractDependentNodes($stemWord_l$)
d_2=extractDependentNodes($stemWord_m$)
for $d_k \in d_1$ **do**
if $d_k \neq d_1$ **then**
$E_j = E_j \cup (d_k, compoundWord)$
for $d_h \in d_2$ **do**
if $d_h \neq d_2$ **then**
$E_j = E_j \cup (d_h, compoundWord)$
G_j=(N_j, E_j)
G_j=removeStopWordNodes(G_j)
return G_j

3.3. User graph

The goal of the user graph is to capture the context of a user in a way that it can determine the stance of a user. Bag of words (BOW) (Dogan & Uysal, 2020; Hakim et al., 2014; Wang & Zhang, 2020) or co-occurrence-based techniques (Bordoloi & Biswas, 2019; Castillo et al., 2015; Rousseau et al., 2015) cannot capture the context in the way user graph does.

There are two naïve ways to create a graph. The first way is to use adjacent words to form a graph, where an edge between two words represents their adjacency. This kind of graph cannot capture the relationship between entities. For example, if we create a graph of the tweet "Kurd people deprived from their water and killed by forces", in this case, adjacent word-based graph cannot form a link between Kurd people (subject) and forces (object). The second way is to create a graph using co-occurring words. This kind of graph may have large cliques lacking useful information. For example, if we add another tweet "We want free Kurdistan", now the co-occurring words from both tweets will have cliques, i.e., every word is connected to every other word for each tweet. This makes it difficult for the machine learning algorithms to learn useful information.

The proposed user graph converts all the user's tweets into a structure like ontologies, where a simple path in the graph goes from the subject word towards the object word while passing through the verb. This means that the verbs help in connecting user's positions, besides that, the verbs near the subject and object words highlight the suppressor and suppressed. The user graph also helps in understanding the context of a user. For example, in the tweet "Kurd are eradicated by the Army", a link from the word Army (subject) to Kurd (object) goes through a verb "eradicate" - this scenario shows that "Kurds" (suppressed) are supposedly "eradicated" by the "Army" (suppressor).

To create a user graph, we use the dependency tree of Spacy API (spacy.io) to create a structure like subject word(s), verbs, object word(s). Spacy extract the grammatical structure of a sentence to find the dependency among words, each word is labeled with a dependency tag. We use these dependencies to create a dependency tree between the words of the tweet. For multiple tweets of a user j, we get multiple trees, which are merged to create a directed user graph $G_j = (N_j, E_j)$.

The graph creation process is shown in Algorithm 1. The process starts by taking input of all the tweets of a user. Next, the algorithm extracts POS dependency (using *extractPOSDependency*) among the words of a tweet. Each word in the POS dependency is represented by a node. We remove redundancy by stemming words using the *stemWord* function that uses Porter stemming. Afterwards, the edges are formed if there is a dependency among words. However, we do not create an edge whenever words have a compound dependency. The compound dependency indicates that the two words should be combined to form a single unit of meaning. Therefore, we first concatenate the compound words using *concatenate* function. Afterwards, we extract the dependent nodes associated with the compound words using *extractDependentNode* function. Later, we adjust the linkage of each dependent nodes. For example, "Great" and "Britain" have a compound dependency, so the algorithm assigns all the edges of the "Great" and "Britain" nodes to the "Great Britain" node. Afterwards, it retains only the "Great Britain" node and deletes others (i.e., "Great" and "Britain" nodes). While merging nodes, we only consider the graph from the current tweet otherwise it will add irrelevant edges. In the last, we remove stopword nodes using *removeStopWordNodes* function.

Stopword removal is performed after creating the user graph of all tweets, otherwise it will result in isolated components. Thus, missing important relationships between words. The proposed user graph structures tweets by linking subject words and object words with verbs, adjectives, and adverbs. This structure helps in understanding the context of a user.

3.3.1. User graph: Case study

We consider a simple example of two tweets by a user for understanding the graph creation process. The tweets are (a) "Forces genocide Kurd people" and (b) "Kurd people deprived from their water and killed by forces.". We extract dependencies among the words of these tweets, as shown in Table 1. The first column of the table represents the current word, the second

Table 1
Dependency parsing for the sample tweets. This dependency is extracted using Spacy API. (a) shows the dependencies for the first tweet and (b) dependencies of the second tweet.

(a)

Current word	Dependency	Parent word
forces	nsubj	genocide
genocide	ROOT	genocide
kurd	compound	people
people	dobj	genocide

(b)

Current word	Dependency	Parent word
kurd	compound	people
people	nsubj	deprived
deprived	ROOT	deprived
from	prep	deprived
their	poss	water
water	pobj	from
and	cc	deprived
killed	conj	deprived
by	agent	killed
forces	pobj	by

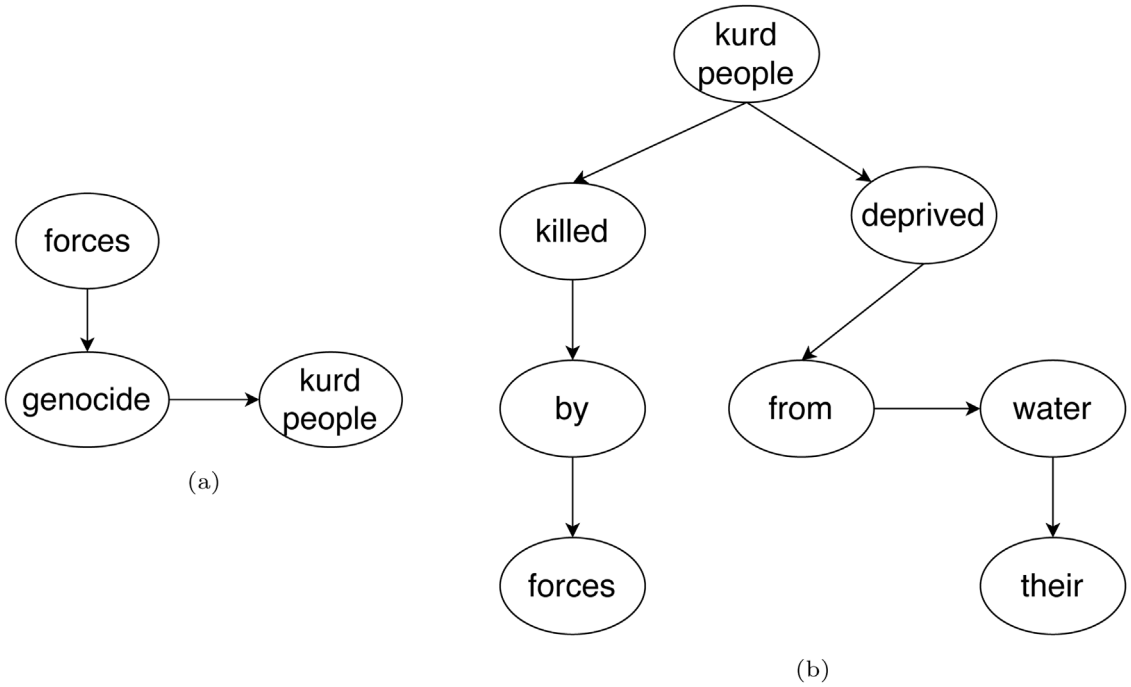

Fig. 2. Dependency tags (see Table 1) are used to create initial trees. (a) Represents the tree created from the first tweet and (b) represents the tree created from the second tweet.

column shows the dependency of the current word and parent word. The third column shows the parent word of the current word. We use these dependencies to create a tree of each tweet as shown in Fig. 2. Both trees are later merged.

Fig. 2 shows the initial tree representation from Table 1, the words “Kurd” and “people” are combined, because they are compound. Afterwards, we stem each node label to remove redundancy. Once a tree is formed from one tweet, we merge the links and new nodes from the dependency tree of the next tweet, thus converting a tree into a graph as shown in Fig. 3. The trees are merged by connecting common words to form a graph. For example, “Kurd people” and “forces” are common words in Fig. 2, due to which trees are merged to form a graph. Next, we delete stopwords after creating a complete graph from all the tweets of a user. Fig. 4 shows an updated graph after removing the stopwords; in this case, we used only one tweet. However, in a real scenario, the graph in Fig. 2 will expand until the system merges all the nodes and edges from all the tweets.

In the user graph, we assume that rebel users may have strong negative verbs (like genocide or atrocities) near the entities, for example, Army, Pakistan, UK, etc. It is evident from Fig. 5 that rebel users frequently use negative words and most of

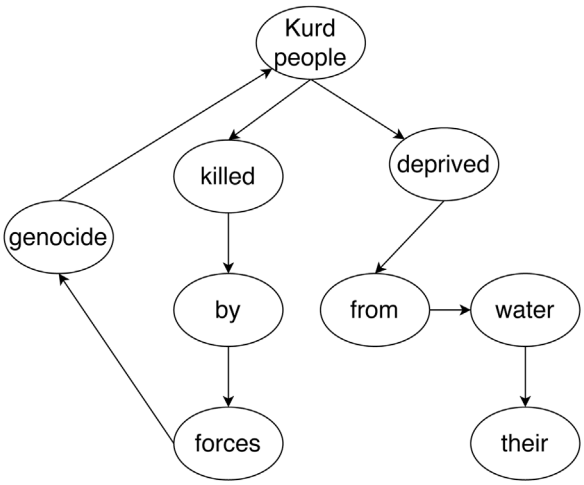

Fig. 3. Graph after merging trees in Fig. 2.

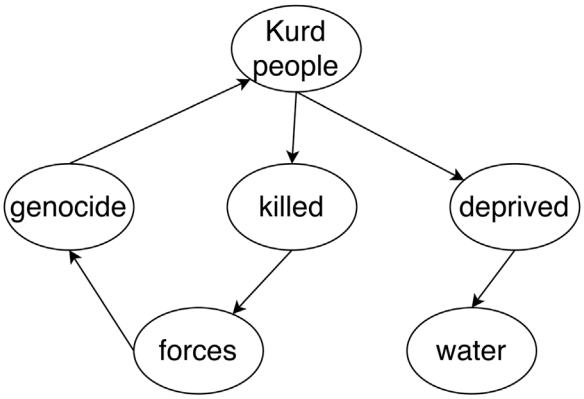

Fig. 4. Updated graph after removing the stopwords from the graph.

these words have a direct link with the government entities. For example, strong verbs (like abduct, attack, force, or terror) are linked to the nouns Army, Pakistan, and Baluchistan, as shown in Fig. 5. This user graph can address questions like what are their demands? who are the claimed victims? and what are their accusations?

On the other hand, counter rebel users use few strong verbs with high centrality, however a few words (like attack, terrorist, Baluchistan, or Pakistan) overlap with the rebel users. In the graph shown in Fig. 6, the counter rebel user is sharing his thoughts on Pakistan efforts on fighting terror. For example, the word "Kill" shows the context that Pakistan must neutralize fighters in their region. On the other hand, the word "Fought" is depicting Pakistan's efforts on fighting terrorism. Overall, this counter rebel user has similar words like a rebel user, therefore it is important to understand the context by looking into the neighboring verbs of various entities. Moreover, the proposed graph can gather the relationship between the long distant words by linking subject and object words with verbs. On the other hand, content features consider the words in isolation, such kind of relationships are not captured.

Apart from the counter rebel users, normal users may not have a large common vocabulary. Therefore, as shown in Fig. 7, the graph has a more positive set of terms. Normal users tend to share their thoughts on general aspects as depicted in Fig. 7, where the graph is showing users' liking towards movies and cars. For example, the link between "Ferrari" and "movies" shows that there might be many movies having Ferrari in them. Generally, normal users tend to have a few strong words with high centrality, which is similar to counter rebel users. Therefore, the challenge for the machine learning algorithms is to separate normal and counter rebel users from rebel users.

3.3.2. Graph embedding

The proposed user graph captures the intention of a user from the history of tweets. This information is then converted into an embedding so that the insights of the user graph are used in machine learning algorithms. Graph embedding is a means to convert a graph into a lower-dimensional vector representation. These lower-dimensional vectors are used in machine learning algorithms to learn patterns within the user graph. In this research, we used graph2vec (Narayanan et al., 2017). The authors use the confined exploration of the neighborhood within the graph to create an embedding. It

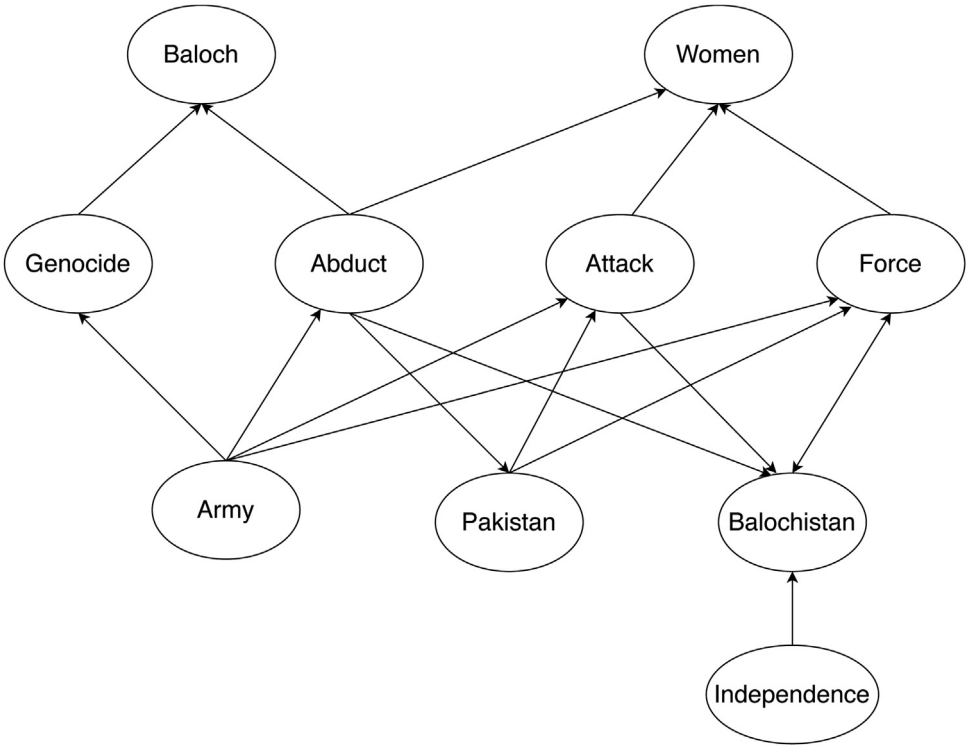

Fig. 5. The subgraph of a rebel user extracted from the proposed user graph. The proposed Algorithm 1 forms relationships among nodes.

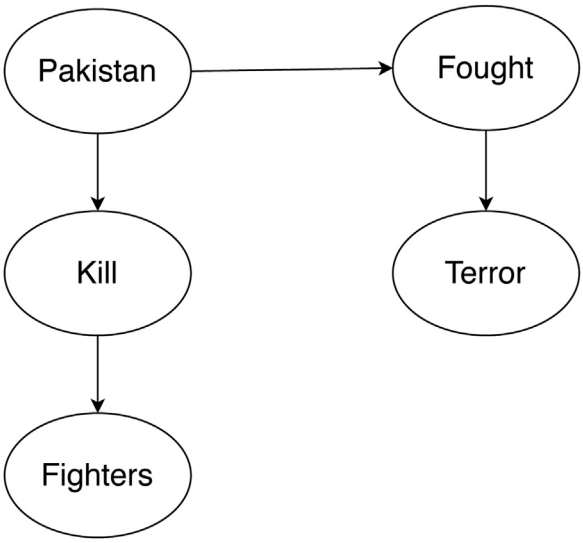

Fig. 6. The subgraph of a counter rebel user extracted from the proposed user graph.

is important to note that this graph embedding approach considers the direction (using neighborhood) of the edge while creating an embedding. This allows us to learn the stance as well as the intention of a user.

3.4. Features selection

Using all the features in the proposed SRI framework may not be useful unless they increase the performance. For this sake, we use the wrapper method to pick the best combination of features (Li et al., 2017). In the wrapper method, we supply all the features to the algorithm, then using machine learning algorithms we pick the best features and their combinations. To combine all the features, we have concatenated the feature vectors of each feature space. In this research, we evaluate

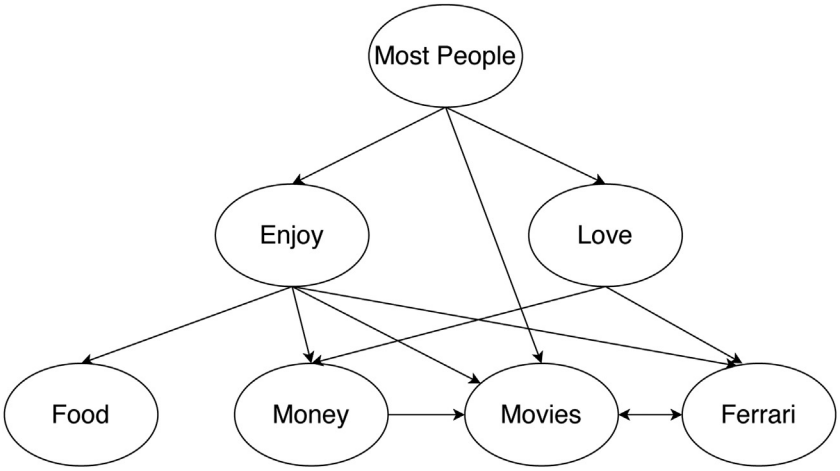

Fig. 7. The subgraph of a normal user extracted from the proposed user graph.

the model for different combinations of features to pick highly-discriminant features. The SRI framework uses key features and their combinations to identify rebel users. The aim of this component is to exclude less discriminant features so that the SRI framework uses useful insights to increase the accuracy.

4. Dataset creation

To create the dataset, we need to first select the countries having rebel movements. First, we extract a sorted list of countries in descending order of the number of incidents since 1970 using the Global Terrorism Database (GTD)[1] . Afterwards, we pick active rebel movements in each country from a Wikipedia article[2] . We remove those countries that have less than six active rebellious movements. Next, we pick top 15 countries, and randomly selected five countries: United Kingdom, Turkey, Syria, Iraq, and Pakistan.

After selecting the countries, we pick those rebellious movements that have an active presence on Twitter, and they have thousands of followers. After filtering, we selected nine rebel movements: Welsh Independence, Kurdistan worker's party (PKK), People's protection unit (YPG), Baluchistan liberation army (BLA), Baloch Student Organization (BSO Azad), Baluchistan Republican Army (BRA), Jeay Sindh Muttahida Mahaz (JSSM), Balawaristan National Front, and Pashtun Tahafuz Movement (PTM); the ideology of each movement is to demand a sovereign state using armed or political means.

Once we select the list of movements, we identified known rebels from the news reports. Afterwards, we manually searched for their Twitter accounts, which served as the seed users. This process led to 15 rebel users. We expanded the list of users using a technique called snowball sampling (Goodman, 1961). In Twitter, lists are used in snowball sampling (Jarwar et al., 2017; Wu, Hofman, Mason, & Watts, 2011). A list contains Twitter users. For the rebel users dataset, we had a total of 11 Twitter lists named as "Extremist", "Free Syria tweets", "Extremist hate groups", "Team Scotland", "Free Balochistan", "Give me freedom Baloch", "Domestic terrorist", "Teapublican terrorist", "Terrorist sympathizer", "Terrorist supporters", and "Known terrorist'. Snowball sampling could not find users of some of the rebellious movements because of their absence in any Twitter lists; in such cases, we used keywords (like abduction, genocide, and others) to manually search rebel users. After searching, we manually classified their profiles by looking into their profiles, pictures, and tweets. We chose those users who had shown affiliation or a positive stance towards a rebel movement.

For counter rebel users, we did not find any Twitter lists. Therefore, we hand-picked each user using the same keywords such as (Abduction, Afrin liberation forces, army, atrocity, genocide, martyr, or force). In case of counter rebel users, we viewed their profiles, and picked those users who had similar attributes (like tweet count and vocabulary), but different stance.

The selection of rebel and counter rebel users is sometimes tricky because we need to understand the history, cultural norms, and recent events to correctly label a user. Therefore, we verified all the users in the dataset from two experts. After getting labels from the experts, we evaluated the agreement between the raters using Cohen's Kappa coefficient. The Kappa value of 0.82 showed a significant agreement between the two raters (McHugh, 2012). The dataset contained only those users for whom the annotators agreed.

[1] https://www.start.umd.edu/gtd/search/
[2] https://en.wikipedia.org/w/index.php?title=List_of_active_rebel_groups&oldid=885545897

Table 2
Statistics of the dataset.

Component	Rebel	Counter rebel	Normal
Number of processed tweets	73,423	94,746	182,085
Average number of processed tweets per user	1049	1373	2985
Number of words	933,025	1,367,066	1,795,214
Average number of words per user	13,329	19,812	29,429
Number of users	70	69	61

For normal users, we aim to pick random users who are neither rebels nor counter rebels. For this sake, we selected a publicly available Twitter dataset[3] consisting of 20,000 normal user accounts. We picked this dataset because the sampling of user accounts was random, which includes many users having normal activities. Apart from this, we skipped the organizational accounts. Finally, we picked a random sample of more than 100 accounts.

In the end, the proposed dataset consisted of 300 users consisting of both rebel, counter rebel, and normal users. We filtered those users by removing those who have less than 10 English tweets. The filtering process results in 70 rebel, 69 counter rebel, and 61 normal users. Table 2 shows the dataset statistics. Finally, we used the Twitter API to extract up to last 3200 tweets of each user. These tweets were gathered and processed in accordance with Twitter's terms and conditions[4] . We removed the non-English tweets, all retweets, and the tweets with less than two words. In total, we extracted a total of 284K tweets, where 73K tweets belonged to rebel users, 94K tweets belonged to counter rebel users, and 182K belonged to normal users.

5. Experiments

In this section, we evaluate the effectiveness of the SRI framework against the two state-of-the-art baselines. We first explain the preprocessing steps. Afterwards, we discuss the baselines. Finally, we discuss the results in detail.

5.1. Data preprocessing

To use the proposed features effectively in the machine learning algorithms, we first remove all the retweets of a user to get his/her own opinion. Furthermore, we replaced the URLs and mentions in the tweets with fixed words (i.e., URL and MENTION). Afterwards, we removed stopwords (in content features) using NLTK English corpus; however, we do not remove negation words (like not or would not) because they are helpful to understand the stance of a user. For example, the tweet saying "I do not hate the UK" shows an opposite perspective after removal of the negation word. In graph embedding, we removed the stopwords towards the end. After removing stopwords, we removed redundant words by performing stemming using Porter stemmer.

5.2. Baseline approaches

To the best of our knowledge there is no publicly available framework for identifying rebel users. In this research, we emphasize on validating whether the proposed framework adds more contextual information for the identification of rebel users. Therefore, we first picked one of the common approaches to represent features in natural language processing, also known as bag of words (BOW) model. BOW model allows conversion of textual information into vector representation. However, such scheme may not be effective in discriminating, for this sake, we use a more effective technique for scoring also known as Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF model helps in evaluating whether the proposed SRI models add more discriminating features.

The problem with TF-IDF is its lack of modeling contextual information. Recently, researchers have used document embedding for capturing the context. Therefore, we picked document embedding as the second baseline, which is among the state-of-the-art approaches to identify the contextual information within a document. By picking document embedding as a baseline, we can evaluate whether the proposed user graph adds more semantics.

5.3. Results and discussions

In this research, we first evaluate the accuracy of the SRI framework using 10-fold cross validation. As mentioned in the Supervised Rebel Identification (SRI) framework, we apply different combination of the proposed features on four machine learning algorithms, i.e., SVM, Random Forest, Gaussian Naive Bayes, and Logistic Regression. First, we only use the content features - the aim is to validate whether the contents alone can produce discriminative features. Second, we only use the user profile features, we assume that the user influence attributes within the user profile might have discriminative attributes.

[3] https://www.kaggle.com/crowdflower/twitter-user-gender-classification
[4] https://developer.twitter.com/en/developer-terms/agreement-and-policy

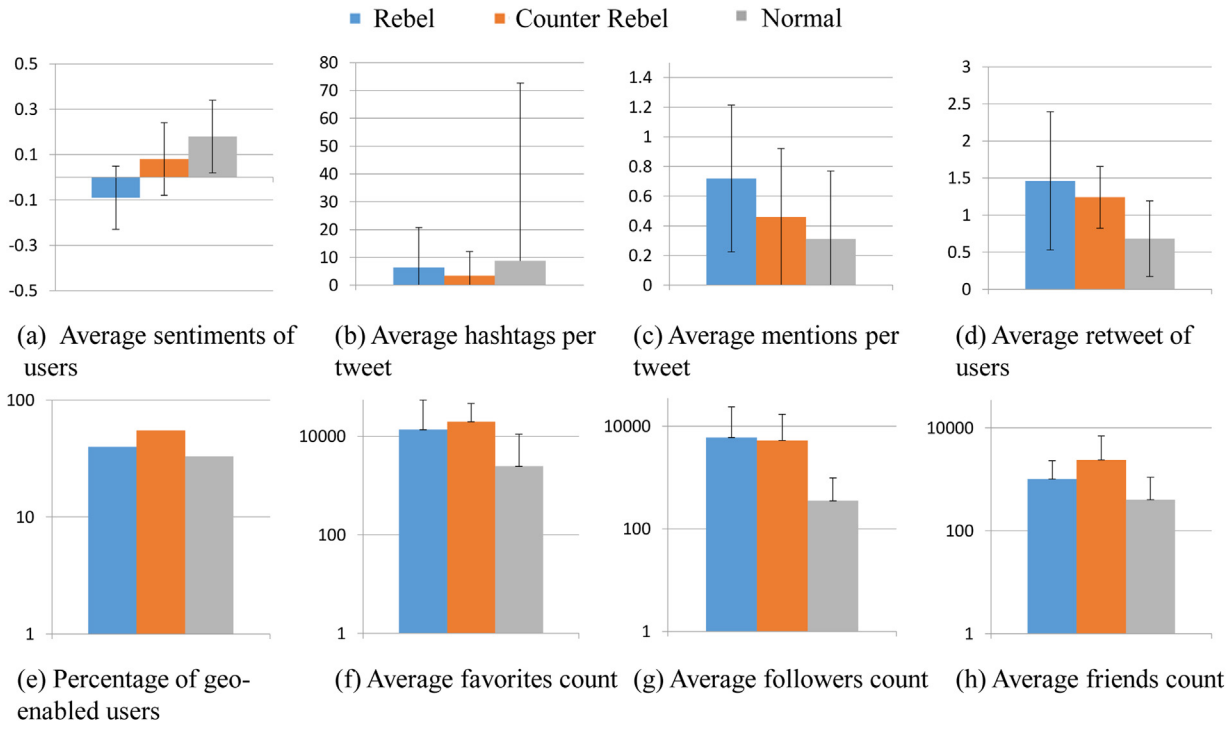

Fig. 8. Analysis of selected features in the dataset. All graphs (a)-(h) except (e) show the averages of various features, (e) shows the percentage. The vertical lines in the graphs show the standard deviations.

Third, we only use the graph embedding, the aim is to add semantics in the state-of-the-art document embedding technique. The fourth feature set combines content and user profile features by merging all the vectors of these feature. This combination allows to add more context to the user profile features. The fifth feature set combines content features and graph embedding vectors into a single vector for each user. Similarly, the sixth feature set merges user profile features and graph embedding to support user profile features with the user's context within the graph embedding. The last feature set combines all the proposed features vectors. In each of the above combinations, the vectors of the features are merged, which increases the vector space for the algorithm to find more patterns. In the proposed SRI framework, we have used wrapper method to identify best combination of features by picking those features that adds more value in classification (Li et al., 2017). The discussion is divided into two parts: the first part discusses results about the model that uses individual feature spaces. The second part discusses performance of the models combining multiple feature spaces.

5.3.1. Single feature space models

As shown in Fig. 9, the models with the single feature set, specifically, content and user profile-based models have lower performance. This means that the model is unable to correctly classify the users using either the content or user profile features. However, the performance of content features is better than the user profile because the vocabulary of rebel and normal users may differ. On the other hand, the vocabulary of the rebel, counter rebel and normal, counter rebel users may overlap with each other, which increases the false positive rate. It is important to note that content features only consider standalone words, which may miss the context. As shown in Fig. 8(a), the sentiments of rebel and counter rebel users are quite similar. However, the sentiments of the normal users are slightly higher than the counter rebel user, thus it further increases the challenge for the classifier to separate the rebel, counter rebel, and normal users. As the sentiments of rebel users are quite close to neutral, therefore, we further investigate the tweets of rebel users and find that the sentiment classifier assigns positive sentiments to the rebel tweets because the words like independence or freedom are considered as positive by the sentiment classifier.

The best performing algorithms (using content features) based on the precision score are the Logistic Regression, SVM, and Gaussian Naïve Bayes. As seen in Fig. 9, the results of content-based algorithms are marginally better than the TF-IDF baseline. However, a noise factor can be seen in Fig. 10, as most of the algorithms under-performed in terms of recall score than the basic TF-IDF baseline. Overall, as per the F1-score in Fig. 11, the LR classifier has the best performance using contents features. In general, the content features have not effectively classified the users, thus it rejects the hypothesis that content, sentiments, and other features together may help identify rebel users.

Before we discuss the result of the user profile features, we first discuss the analysis of the features within the user profiles. In general, the assumption behind user profile features is that majority of the rebel users use Twitter for their

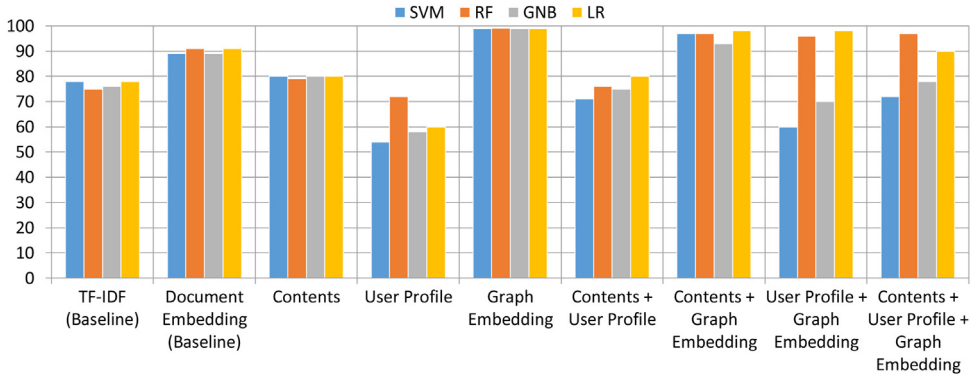

Fig. 9. Precision scores of different approaches for various features.

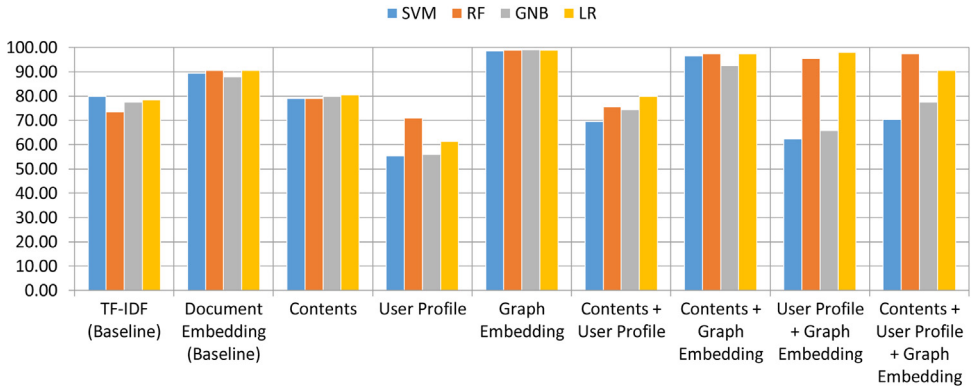

Fig. 10. Recall score of different approaches on various features.

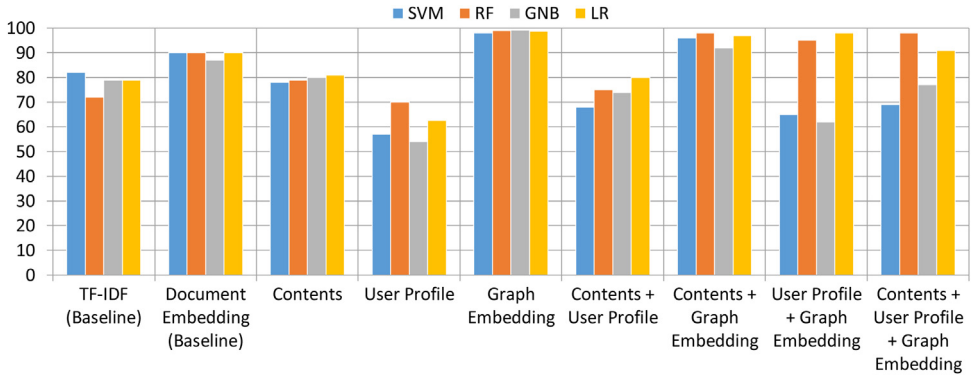

Fig. 11. F1-score of different approaches on various features.

cause, therefore their profile descriptions will contain more mentions and hashtags in comparison to other users. However, Figs. 8(b) and (c) show that average mention count and average hashtag count are also quite similar between the rebel and counter rebel users. However, there is a slight difference in the average hashtags and mentions count of normal users, but the difference is nominal. Similarly, Fig. 8(d) shows that the normal users get more retweets, whereas rebel users gain a smaller number of retweets. Another aspect shows that the counter rebel users may get more favorites as shown in Fig. 8(f). Fig. 8(g) shows the same phenomenon, where rebel and counter rebel users have a similar following, whereas normal users have slightly fewer followers. Generally, rebel and counter rebel may have an ideology, therefore, people who believe in that ideology may follow them. On the other hand, normal users may not promote their ideology in their profiles. Finally, as shown in Fig. 8(h), the friend count is quite similar to follower count except for the fact that the rebel users do not gain the same number of friends than normal or counter rebel users. This might be due to the reason that many journalists or researchers follow them.

Generally, most of the profile features of the rebel and counter rebel users and normal and counter rebel users are quite similar, which makes the machine learning task difficult. As shown in Figs. 9–11, the precision, recall, and f1-score of the user profile features are even lower than the content features. This is because user profiles do not have discriminating features, as discussed above, most of the features in the user profile have similar scores. As per the user profile features, the RF-based algorithms have better performance in comparison to others because they are less susceptible to noise. The observed results helped in answering that the user profile features may have similar features, thus the algorithms relying on them may not help the classification.

Afterwards, we evaluate the performance of the proposed user graph (graph embedding). The results of the precision in Fig. 9 shows that all the algorithms have outperformed the baselines approaches. The precision score is above 99% irrespective of the classifier, meaning that the proposed user graph can accurately separate the users.

Interestingly, as shown in Fig. 10, the recall scores from the proposed user graph and its embedding are near to perfect. This means that the user graph can correctly identify the context, thus having a higher coverage. Besides that, Fig. 11 shows that the graph embedding features have outperformed all the baselines with an F1-score more than 98%. There is a significant improvement in the F1-score of graph embedding features over document embedding. The best algorithm with graph embedding features is the RF, although the difference with other algorithms is nominal. In the context of the user graph, we say that the proposed user graph helps all the tested machine learning algorithm to learn the contextual information. Moreover, the proposed user graph creation mechanism accurately identifies the perspective or stance of a user even if it contains overlapping terms in all classes, i.e., rebel, counter rebel, and normal. For example, as shown in Fig. 6, the direction and relationship of the words in the user graph gives an indication about the context of the users. On the other hand, content features do not form the relationships among words, which is essential to identify the context of a user. Therefore, the performance of content features is lower than the graph embedding.

5.3.2. Multiple feature space models

In this research, we hypothesize that multiple features together may help in increasing the accuracy. Similarly, we also hypothesize that user profile features can be effective in conjunction with other features. However, combining the content and user profile features did not result in a better performance. As combining these features may increase noise. This can be observed in the Fig. 11, where all the algorithms have under-performed even the content-based algorithms, let alone the baselines.

In the same manner, most of the models with graph embedding as one of the features have mostly outperformed the baselines. Naturally, this improvement is based on the user graph because it helps in accurately identifying the context of a user. Interestingly, most of these algorithms (with graph embedding) have higher recall scores as shown in Fig. 10. However, these models perform lower than the graph embedding features alone. This is because of the lack of context in both content and user profile features. Moreover, in Fig. 11, mostly the models with one feature as content or user profile and the other feature as graph embedding have also shown improvement over the document embedding and TF-IDF baselines.

The results of the content and graph embedding features show significant improvements over baselines. However, they have slightly lower performance than the graph embedding based single feature model. On the other hand, the results of the user profile and graph embedding features with the GNB and SVM algorithms have underperformed the baselines. Generally, SVM and GNB are more sensitive to noise (Atla, Tada, Sheng, & Singireddy, 2011; Huang, Shi, & Suykens, 2014; Shieh & Kamm, 2009). As discussed earlier, user profile features have more noise, so the results reject the hypothesis that the user profile feature can be helpful if merged with other features.

Finally, we use all the features together to see the performance of the different algorithms. All the algorithms outperform the baselines in terms of F1-score (as shown in Fig. 11) except the SVM algorithm. However, this model slightly underperformed the model with single graph embedding based features. Overall, the proposed SRI model with the user graph and its embedding performs the best in identifying the rebel users. This shows that the user graph is generic enough to identify context regardless of the region or culture. As in the results, the user graph even with diverse set of users has been able to achieve around perfect accuracy. This shows that the user graph can help algorithms accurately separate users regardless of the region or culture. Moreover, the proposed user graph can manage regional and cultural aspects because it relies on the hypothesis that rebel users mostly target government entities. Therefore, a rebel user from any region or culture targets government entities to gain more acceptability in his community. Thus, the user graph and its embedding will use capture such relationships to identify rebel users.

Graph embedding has the best results, but when it is combined with other features, the accuracy decreases. This means that the proposed user graph has the most impact on the accuracy because it creates a detailed picture of a user by semantically structuring all the tweets of a user in a graph. Furthermore, the direction of links in the proposed user graph also determines the context of a user. The user graph assumes that it may have strong verbs nearby the target entities with high degree centrality. If we consider the target entities to be “Baluchistan” and “Army” in Fig. 5, now the link between “Army” and strong verbs (like genocide, abduction) shows the reasons, whereas the link between “independence” and “Baluchistan” shows the stance and ideology. Also, the degree centrality of the target entities and strong verbs is high, thus showing their user’s dedication towards the cause. On the contrary, counter rebel users may not have same level of dedication, so the degree centrality in their user graph is low. Besides that, they use a mixture of positive and negative verbs that mostly depict criticism or concern on a certain issue. As a result, a deeper traversal into the neighborhood of the user graph helps in understanding the stance or ideology of a user. The assumption that the rebel users will use more strong verbs near the

entities is also true. By looking into the user graph, we can identify the class of a user without even looking at his profile. Therefore, we convert this user graph into a lower dimension embedding, which results in highly accurate classification.

6. Conclusions

We propose a novel SRI framework to identify rebel users. To evaluate the framework, we also propose the first multicultural and multiregional dataset of nine rebel movements belonging to five countries. The results indicate improvements over the state-of-the-art baselines. In the SRI framework, we propose a total of fifteen features, among them, we propose a novel mechanism to structure the user tweets in a user graph. The user graph converts the tweets of a user into an ontology like structure, where the subject and object words are linked with predicates. The results indicate that the proposed user graph captures more semantics regardless of the region or culture. Even with a diverse set of users, the proposed user graph can correctly identify rebel users with high accuracy. In this research, we found out that normal sentiment classification techniques may not identify the correct sentiments because of the complexity in the statements of rebel users.

This research has potential benefits from various aspects, first, the proposed framework can help identify other kind of users (like scientists, journalist, and experts) by capturing their contexts using the proposed user graph. Second, the proposed framework will provide a benchmark for future research related to rebels. Third, it will help sociologists and the government to better understand behavior of rebel users and how to address their concerns. Fourth, the proposed framework can help social media platforms like Twitter to reduce the spread of hatred by identifying users who spread it.

Authors' contribution

Muhammad Ali Masood: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.

Rabeeh Ayaz Abbasi: Conceived and designed the analysis; Wrote the paper.

References

Agarwal, S., & Sureka, A. (2015). A topical crawler for uncovering hidden communities of extremist micro-bloggers on tumblr. *5th workshop on making sense of microposts (MICROPOSTS)*, 1–2.

Alvari, H., Sarkar, S., & Shakarian, P. (2019). Detection of violent extremists in social media. *2019 2nd international conference on data intelligence and security (ICDIS)*, 43–47.

Atla, A., Tada, R., Sheng, V., & Singireddy, N. (2011). Sensitivity of different machine learning algorithms to noise. *Journal of Computing Sciences in Colleges*, *26*(5), 96–103.

Badawy, A., & Ferrara, E. (2018). The rise of jihadist propaganda on social networks. *Journal of Computational Social Science*, *1*(2), 453–470.

BBC. (2019). *Twitter 'confuses' iyad el-baghdadi with islamic state leader*. (last accessed, Nov 6, 2020;) URL https://www.bbc.com/news/world-35210527

Bermingham, A., Conway, M., McInerney, L., O'Hare, N., & Smeaton, A. F. (2009). Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. *Social network analysis and mining, 2009. ASONAM'09. international conference on advances in*, 231–236.

Bordoloi, M., & Biswas, S. K. (2019). Graph-based sentiment analysis model for e-commerce websites' data. In P. K. Mallick, V. E. Balas, A. K. Bhoi, & A. F. Zobaa (Eds.), *Cognitive informatics and soft computing* (pp. 453–462). Singapore: Springer Singapore.

Castillo, E., Cervantes, O., Vilarino, D., Báez, D., & Sánchez, A. (2015). Udlap: Sentiment analysis using a graph-based representation. *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 556–560.

Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015). Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. *Proceedings of the 16th annual international conference on digital government research*, 239–249.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, *37*(4), 50–56.

Choi, D., Ko, B., Kim, H., & Kim, P. (2014). Text analysis for detecting terrorism-related articles on the web. *Journal of Network and Computer Applications*, *38*, 16–21.

Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016). A novel text mining approach based on tf-idf and support vector machine for news classification. *2016 ieee international conference on engineering and technology (ICETECH)*, 112–116.

Debnath, S., Das, D., & Das, B. (2017). Identifying terrorist index (t+) for ranking homogeneous twitter users and groups by employing citation parameters and vulnerability lexicon. *International conference on mining intelligence and knowledge exploration*, 391–401.

Dogan, T., & Uysal, A. K. (2020). A novel term weighting scheme for text classification: Tf-mono. *Journal of Informetrics*, *14*(4), 101076. http://dx.doi.org/10.1016/j.joi.2020.101076

Fernandez, M., & Alani, H. (2018). Contextual semantics for radicalisation detection on twitter. *semantic web for social good workshop (SW4SG) at international semantic web conference 2018*, 1–14.

Fernandez, M., Asif, M., & Alani, H. (2018). Understanding the roots of radicalisation on twitter. *Proceedings of the 10th acm conference on web science*, 1–10.

Gialampoukidis, I., Kalpakis, G., Tsikrika, T., Papadopoulos, S., Vrochidis, S., & Kompatsiaris, I. (2017). Detection of terrorism-related twitter communities using centrality scores. *Proceedings of the 2nd international workshop on multimedia forensics and security*, 21–25.

Goodman, L. A. (1961). *Snowball sampling, The annals of mathematical statistics*. pp. 148–170.

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014). Automated document classification for news article in bahasa indonesia based on term frequency inverse document frequency (tf-idf) approach. *2014 6th international conference on information technology and electrical engineering (ICITEE)*, 1–4.

Haneef, F., Abbasi, R. A., Sindhu, M. A., Khattak, A. S., Noor, M. N., Aljohani, N. R., Daud, A., & Arafat, S. (2020). Using network science to understand the link between subjects and professions. *Computers in Human Behavior*, *106*.

Hartung, M., Klinger, R., Schmidtke, F., & Vogel, L. (2017). Identifying right-wing extremism in german twitter profiles: A classification approach. *International conference on applications of natural language to information systems*, 320–325.

Huang, X., Shi, L., & Suykens, J. A. K. (2014). Support vector machine classifier with pinball loss. *Ieee transactions on pattern analysis and machine intelligence*, *36*(5), 984–997.

Husslage, B., Borm, P., Burg, T., Hamers, H., & Lindelauf, R. (2015). Ranking terrorists in networks: A sensitivity analysis of al qaeda's 9/11 attack. *Social Networks*, *42*, 1–7.

Jarwar, M. A., Abbasi, R. A., Mushtaq, M., Maqbool, O., Aljohani, N. R., Daud, A., Alowibdi, J. S., Cano, J., García, S., & Chong, I. (2017). CommuniMents: A Framework for Detecting Community Based Sentiments for Events. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *13*(2), 87–108.
Li, D., Ding, Y., Shuai, X., Bollen, J., Tang, J., Chen, S., Zhu, J., & Rocha, G. (2012). Adding community and dynamic to topic models. *Journal of Informetrics*, *6*(2), 237–253.
Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*, *50*(6).
Masood, M. A., Abbasi, R. A., & Wee Keong, N. (2020). Context-aware sliding window for sentiment classification. *IEEE Access*, *8*, 4870–4884.
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276–282.
Mueller, J. A. (2014). *International norm echoing in rebel groups: The cases of the kosovo liberation army and the liberation tigers of tamil eelam*. New York, USA: City University of New York. Ph.D. thesis.
Nam, M., Lee, E., & Shin, J. (2015). A method for user sentiment classification using instagram hashtags. *Journal of Korea Multimedia Society*, *18*(11), 1391–1399.
Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017). graph2vec: Learning distributed representations of graphs. In *Proceedings of the 13th international workshop on mining and learning with graphs (MLG)*. pp. 1–8. Halifax, Nova Scotia, Canada: MLG.
Ovelgonne, M., Kang, C., Sawant, A., & Subrahmanian, V. (2012). Covertness centrality in networks. *Advances in social networks analysis and mining (ASONAM), 2012 ieee/acm international conference on*, 863–870.
Pan, S. J., Ni, X., Sun, J.-T., Yang, Q., & Chen, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on world wide web, WWW '10*. pp. 751–760. New York, NY, USA: Association for Computing Machinery.
Reynolds, S. C., & Hafez, M. M. (2017). Social network analysis of german foreign fighters in Syria and Iraq. *Terrorism and Political Violence*, 1–26.
Richey, M. K., & Binz, M. (2015). Open source collection methods for identifying radical extremists using social media. *International Journal of Intelligence and CounterIntelligence*, *28*(2), 347–364.
Rousseau, F., Kiagias, E., & Vazirgiannis, M. (2015). Text categorization as a graph classification problem. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Vol. 1: long papers), Vol. 1*, 1702–1712.
Rowe, M., & Saif, H. (2016). Mining pro-isis radicalisation signals from social media users. *Proceedings of the 10th international conference on web and social media*, 329–338.
Saif, H., Fernández, M., Rowe, M., & Alani, H. (2016). On the role of semantics for detecting pro-isis stances on social media. *CEUR workshop proceedings, Vol. 1690*, 1–4.
Scanlon, J. R., & Gerber, M. S. (2014). Automatic detection of cyber-recruitment by violent extremists. *Security Informatics*, *3*(1), 5.
Schneider, N. K. (2015). *Isis and social media: The combatant commander's guide to countering isis's social media campaign, Tech. rep., DTIC Document*. 8725 John J. Kingman Road, Fort Belvoir, VA 22060-6218, USA: Defense Technical Information Center.
Seyednezhad, S. M. M., Fede, H., Herrera, I., & Menezes, R. (2018). Emoji-word network analysis: Sentiments and semantics. *The thirty-first international flairs conference*, 271–274.
Shieh, A. D., & Kamm, D. F. (2009). Ensembles of one class support vector machines. In J. A. Benediktsson, J. Kittler, & F. Roli (Eds.), *Multiple classifier systems* (pp. 181–190). Berlin, Heidelberg: Springer Berlin Heidelberg.
Sindhwani, V., & Melville, P. (2008). Document-word co-regularization for semi-supervised sentiment analysis. *2008 Eighth ieee international conference on data mining*, 1025–1030.
Skillicorn, D. (2010). Applying interestingness measures to ansar forum texts. *ACM SIGKDD workshop on intelligence and security informatics*, 7.
Sureka, A., Kumaraguru, P., Goyal, A., & Chhabra, S. (2010). Mining youtube to discover extremist videos, users and hidden communities. *Asia information retrieval symposium*, 13–24.
Sureka, A. (2012). *140 characters of @ hate and# protest, Tech. rep.* IIIT Delhi, India: IIITD-TR-2011-007.
J. Tinnes, The art of searching: how to find terrorism literature in the digital age, Perspectives on Terrorism 7 (4).
Wadhwa, P., & Bhatia, M. (2013). Tracking on-line radicalization using investigative data mining. In *Communications (NCC), 2013 National Conference on*. pp. 1–5. IEEE.
Wadhwa, P., & Bhatia, M. (2014). Classification of radical messages in twitter using security associations. In *Case studies in secure computing: Achievements and trends*. pp. 273–294.
Wadhwa, P., & Bhatia, M. (2015). An approach for dynamic identification of online radicalization in social networks. *Cybernetics and Systems*, *46*(8), 641–665.
Wadhwa, P., & Bhatia, M. (2016). New metrics for dynamic analysis of online radicalization. *Journal of Applied Security Research*, *11*(2), 166–184.
Wang, S. I., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. *Proceedings of the 50th annual meeting of the association for computational linguistics (Vol. 2: Short Papers)*, 90–94.
Wang, Y., & Zhang, C. (2020). Using the full-text content of academic articles to identify and evaluate algorithm entities in the domain of natural language processing. *Journal of Informetrics*, *14*(4), 101091. http://dx.doi.org/10.1016/j.joi.2020.101091
Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th acm international conference on information and knowledge management, CIKM '11*. pp. 1031–1040. New York, NY, USA: Association for Computing Machinery.
Wang, G., Wang, B., Wang, T., Nika, A., Zheng, H., & Zhao, B. Y. (2014). Whispers in the dark: Analysis of an anonymous social network. In *Proceedings of the 2014 conference on internet measurement conference*. pp. 137–150. ACM.
Wei, Y., & Singh, L. (2017). Using network flows to identify users sharing extremist content on social media. In *Pacific-Asia conference on knowledge discovery and data mining*. pp. 330–342. Springer.
Wei, Y., Singh, L., & Martin, S. (2016). Identification of extremism on twitter. In *Proceedings of the 2016 ieee/acm international conference on advances in social networks analysis and mining*. pp. 1251–1255. IEEE Press.
Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web*. pp. 705–714. ACM.
Xu, J., Lu, T.-C., et al. (2016). Automated classification of extremist twitter accounts using content-based and network-based features. In *2016 IEEE international conference on big data (big data)*. pp. 2545–2549. IEEE.